

IPRO 311 – US Holocaust Memorial Museum Archives

- Jason Soo, Computer Science, 2009
- Yacin Nadji, Computer Science, 2009
- Justin Moses, Computer Science, 2009
- Jane Ji, Applied Mathematics, 2008
- Jeff Rowland, ITM, 2007
- Erica Scott, ITM, 2006
- Mark Haase, Chemical Engineering, 2007

Searching the USHMM

- Previous IPROs
 - Developed Mediator
 - Improved general search with relevance feedback (Phase II)
- Immense collection of documents
- Manual Searches only
 - Time consuming
 - Complex
- There are better ways

Better Ways

- Phase I: Locality Search
- Phase II: Search Optimization (Relevance Feedback)
- Testing

Phase I: Locality Search

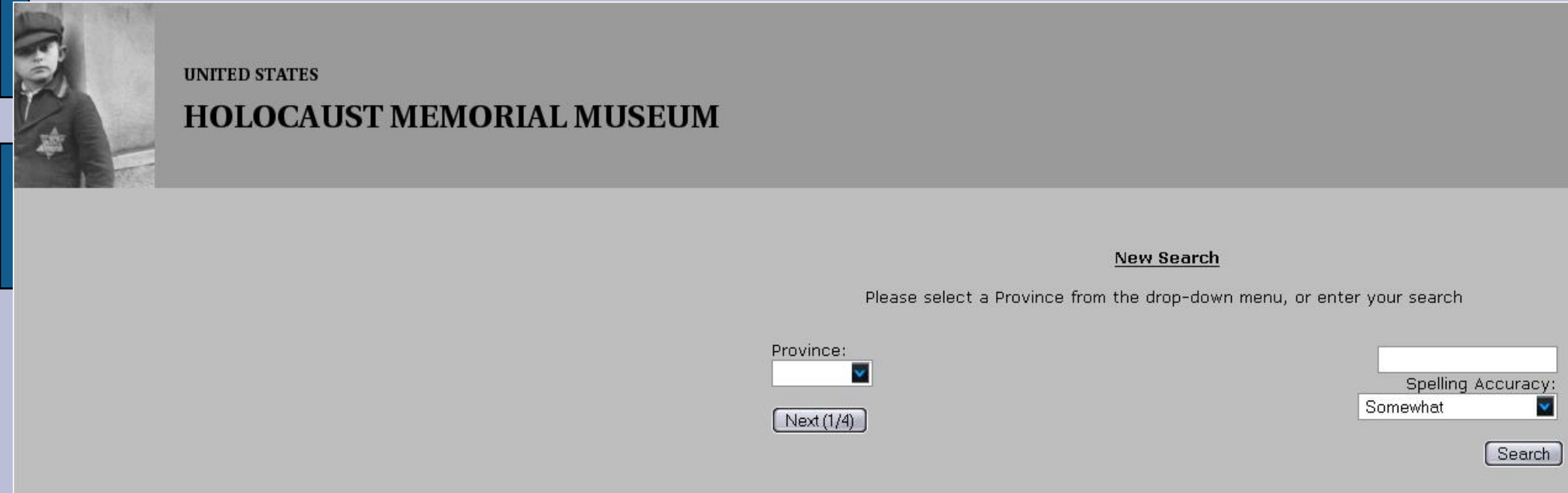
The Facts

- At least one person will call the USHMM looking for answers and they will be turned away.
- Over 45,000 slides plus stacks of papers to look through to find answers.
- We helped change that.

What & Why?

- We developed an online program with a user interface.
- This was done the help the United States Holocaust Memorial Museum, and is scheduled to be in and used by the Holocaust Museum later this year.

How We Helped



The screenshot shows the top section of a website for the Holocaust Memorial Museum. On the left is a small black and white photograph of a young boy wearing a Star of David. To the right of the photo, the text reads "UNITED STATES" and "HOLOCAUST MEMORIAL MUSEUM". Below this is a search interface with the heading "New Search". A prompt asks the user to "Please select a Province from the drop-down menu, or enter your search". There are two drop-down menus: one for "Province:" and one for "Spelling Accuracy:" which is currently set to "Somewhat". There is also a search input field, a "Next (1/4)" button, and a "Search" button.

- Developed an online searcher – Currently running on real Holocaust data.
- Was written intelligently in intelligent languages so that it can be run by almost any server.

First Way To Find Answers Selection Process



UNITED STATES

HOLOCAUST MEMORIAL MUSEUM

[New Search](#)

Please select a Province from the drop-down menu, or enter your search

Province:

Spelling Accuracy:
Somewhat



First Way To Find Answers

Selection Process

Province:

A screenshot of a web form. At the top, the text "Province:" is displayed. Below it is a white rectangular input field with a dotted border. To the right of this field is a blue square button with a white downward-pointing arrow. Below the input field and button is a dropdown menu. The menu is currently open, showing a solid blue bar at the top and a white bar below it containing the text "Slovakia". The dropdown menu has a slight 3D effect with a grey shadow at the bottom.

First Way To Find Answers

Selection Process

You have selected: Slovakia,

District:

Banska Bystrica

Banovce nad Bebravou

Banska Bystrica

Banska Stiavnica

Bardiov

Bracovce, Falkusovce, Racanov, Laskovce, Lozin, Trhoviste

Bratislava

Brezno nad Hronom

Cadca

Dobsina

Dolny Kubin

Gelnica

Giraltovce

Hlohovec

Humenne

Ilava

Kezmarok

Krupina

Kysucke Nove Mesto


Lesne, Moraviany, Nizny Hrusov, Suche

First Way To Find Answers

Selection Process

You have selected: Slovakia, -> Banska Bystrica,

Municipality:



- Banska Bystrica, Horna Micina, Lubietova, Selce
- Slovenska Lupca, Radvan nad Hronom
- Stare Hory, Sv. Ondrej n. Hr.

First Way To Find Answers Selection Process

Slovak Archives - Jewish Census 1942 townships and communities, Reel 1 Box 182 corrected,

[Search again](#)

Second Way To Find Answers

Searching Process



UNITED STATES

HOLOCAUST MEMORIAL MUSEUM

New s

Please select a Province from the drop-down, or enter your search

Province:

Next (1/4)

Spelling Accuracy:
Somewhat

Search

(c) ...

Second Way To Find Answers

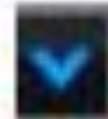
Searching Process

Search

Slovakia

How correct do you believe
you spelled the word?

Somewhat



Search

Second Way To Find Answers

Searching Process

Start Over

We found the province Slovakia,

You may change your search settings if you couldn't find your word.

Search

How correct do you believe you spelled the word?



Second Way To Find Answers

Searching Process

You have selected: Slovakia,

District:

- Banovce nad Bebravou
- Banska Bystrica
- Banska Stiavnica
- Bardiov
- Bracovce, Falkusovce, Racanov , Laskovce , Lozin, Trhoviste
- Bratislava
- Brezno nad Hronom
- Cadca
- Dobsina
- Dolny Kubin
- Gelnica
- Giraltovce
- Hlohovec
- Humenne
- Ilava
- Kezmarok
- Krupina
- Kysucke Nove Mesto
- Lesne, Moraviany, Nizny Hrusov, Suche

Second Way To Find Answers

Searching Process

Search

Slavakia

How correct do you believe
you spelled the word?

Somewhat



Search

Second Way To Find Answers

Searching Process

Start Over

Nothing in our database matches Slavakia
Trying to Help...

- Province Slovakia, is the only close match we found.

You may change your search settings if you couldn't find your word.

Search

How correct do you believe you spelled the word?



Second Way To Find Answers

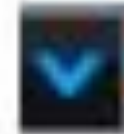
Searching Process

Search

Slavaka

How correct do you believe
you spelled the word?

Somewhat



Search

Second Way To Find Answers

Searching Process

Start Over

Nothing in our database matches Slavaka
Trying to Help...

- Province Slovakia, is 67% is likely to be your choice.
- Municipality Slatina nad Bebravou, is 33% is likely to be your choice.

You may change your search settings if you couldn't find your word.

Search

How correct do you believe you spelled the word?



Statistics Of Our Searcher

- Over 75% of the time the correct word of a misspelled word was found.
- Over 80% of the time the correct word was ranked the highest.
- With two or more words, the correct words were ranked highest over 90% of the time.



UNITED STATES

HOLOCAUST MEMORIAL MUSEUM

New Search

Please select a Province from the drop-down menu, or enter your search

Province:

Next (1/4)

Spelling Accuracy:

Somewhat

Search

Phase II: Optimizing Search

Phase II: Optimizing Search

IPRO 311

- Why does search need to be optimized?
- What did IPRO 311 do to optimize search?
- How can this improve the United States Holocaust Memorial Museum?

Why does search need to be optimized?

- Search is not perfect



The screenshot shows a Google search interface. At the top left is the Google logo. To its right are navigation links: [Web](#), [Images](#), [Video](#), [News](#), [Maps](#), and [more »](#). Below these is a search input box containing the text "three best restaurants in chicago". To the right of the input box is a "Search" button and a link to [Advanced Search Preferences](#).

Web

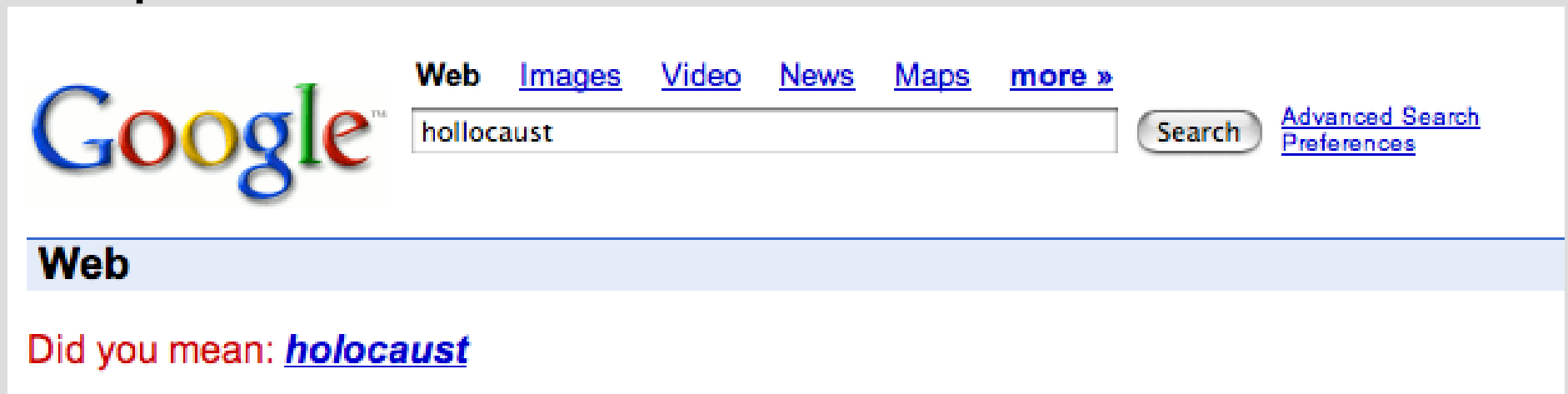
[Map of 3 Best Restaurants In, Chicago, IL](#)
 [Google Maps](#) - [Yahoo! Maps](#) - [MapQuest](#)

[Chicago's Best Restaurants](#)
Chicago's Best Restaurants. Three of Chicago's restaurants are listed in 'America's Best 50 Restaurants' guide.
[chicago.about.com/cs/restaurants/a/top_restaurants.htm](#) - 27k - [Cached](#) - [Similar pages](#)

[Top Ten Chicago Restaurants - Dining - Chicago Restaurants](#)
Gourmet magazine named this Chicago restaurant one of "The Best 50 Restaurants in the United States" in 2001 (it was one of only three Chicago restaurants ...
[chicago.about.com/od/fooddrink/tp/070606_dining.htm](#) - 32k - [Cached](#) - [Similar pages](#)

Why does search need to be optimized?

- Most major search engines already offer optimizations



- While this is helpful, it is not enough

Why does search need to be optimized?

- USHMM search, in particular, is very poor
 - Does not include even simple Google optimizations
- USHMM contains *very important information*
 - *Great resource*
 - *Poor search implementation*
 - *If you cannot retrieve the information, is it useful?*

What did IPRO 311 do to optimize search?

- Fully functional Java-based search assistant
 - Java => multi-platform
- How does it work?
 - Relevance Feedback
- Benefits
 - Improves search
 - Entirely transparent to user

What did IPRO 311 do to optimize search?

- Brainstormed/Implemented 10 functions
 - 1.) Top Terms
 - 2.) Common Terms (Common terms across all documents)
 - 3.) Format (HTML)
 - 4.) Last Results
 - 5.) Names

international Holocaust denial movement: Arthur Butz (USA), Bradley Smith (USA), Fred Leuchter (USA), Ernst Zundel (Germany, Canada), David Irving (Great Britain), Robert Faurisson (France), Ahmed Rami (Sweden), Carlo Mattogno (Italy).

What did IPRO 311 do to optimize search?

Cont.

- 6.) Places
- 7.) Proper Noun
- 8.) Plural
- 9.) Sentences
- 10.) Year

They hold pseudo-scholarly conventions. The Institute for Historical Review in California, a center of the Holocaust denial movement in the United States, has until recently published a pseudo-scholarly

their occupation of Germany in 1945 and the harsh persecution of Nazi defendants. Holocaust deniers also insist that Jews needed the

Example Query

```
ynadji@zchor ~ $ queryushmm "anne"  
java -Dgate.home=/usr/local/ushmm/gate holocaustSearch.Search "anne"  
Number of actual words: 5737  
Adding to query: +frank+diary+
```


Example Query

◀ | UNITED STATES HOLOCAUST MEMORIAL MUSEUM | ▶

Web Site Search Results

[Advanced Search](#)

Search

Narrow your search by choosing

60 of 60 documents returned for the query **anne**.

[Prior](#)

[Next](#)

Documents 1 to 25:

Example Query

◀|UNITED STATES HOLOCAUST MEMORIAL MUSEUM||

Web Site Search Results

[Advanced Search](#)

Search

Narrow your search by choosing ▶

5 of 5 documents returned for the query **anne frank diary**.

What did IPRO 311 do to optimize search?

- Did our search assistant help?
 - Yes, but how do we know?
 - Testing Phase
 - Showed where improvements were being made, and where they were not
 - Caught bugs in the code
 - Forced deadlines to ensure a streamlined, efficient product

How can this improve the USHMM?

- *“Those who cannot remember the past, are condemned to repeat it” - George Santayana (US Philosopher/Poet)*
- *The current conflict in the Darfur region of Sudan (labeled as genocide) is currently spreading to neighboring country Chad (MSNBC Nov. 13th 2006)*
- *Without adequate ways of search, knowledge is lost*

Testing Phase

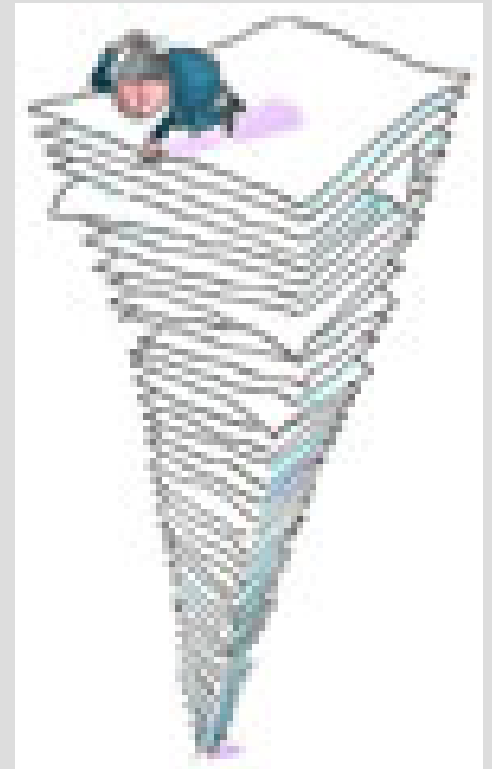
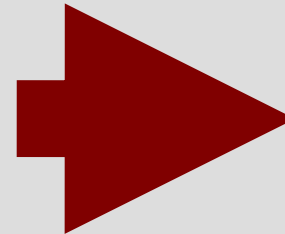
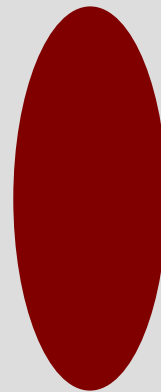
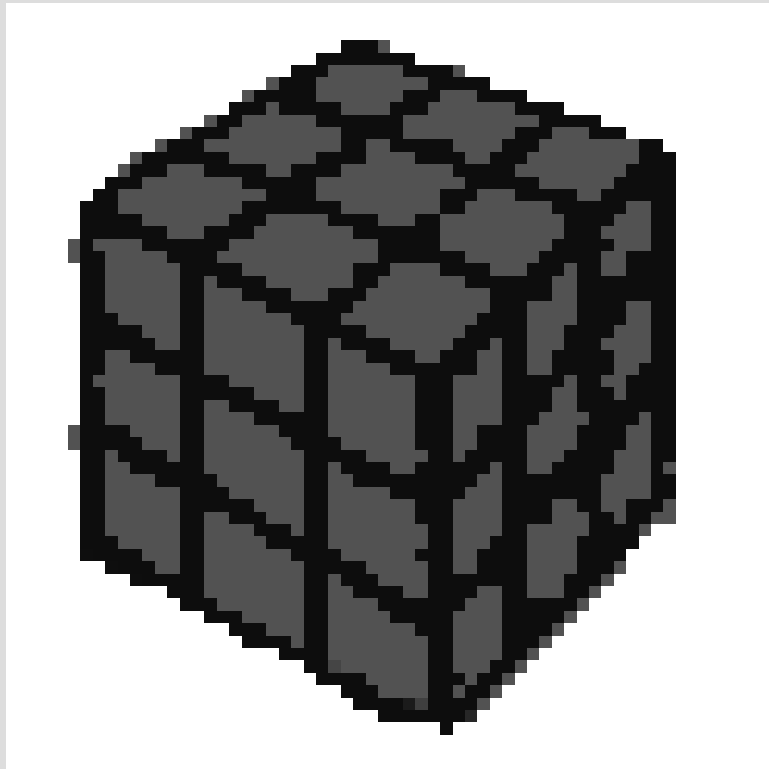
Testing Process

- Tested both Locality Search tool and Search Optimization functions
 - Required partnership and cooperation between all members of the IPRO



Black Box Test Method

- Black Box Test Method
 - Focuses on the output of application being tested



Locality Search Tool Testing

- Tested tools ability to handle requests by multiple users simultaneously
 - Results: The tool is able to handle “system overloads” without distorting data
 - Currently in the process of being deployed to USHMM

Search Improvement Testing

- Developed and executed test cases based on top 200 terms most frequently searched on USHMM site in past year
- Compared output of site with and without enhancements
 - Better
 - Same
 - Worse

Search Improvement Results

- 5% improvement in site functionality
- 0.1% improvement in Google AdSense searches earned an additional \$100 million in revenue



Closing Remarks

Closing Remarks

"The things I saw beggar description...the visual evidence and the verbal testimony of starvation, cruelty and bestiality were so overpowering. I made the visit deliberately, in order to be in a position to give first hand evidence of these things if ever, in the future, there develops a tendency to charge these allegations to propaganda."

– Dwight D Eisenhower, Ohrdruf Camp, April 15, 1945.